Discrete Applied Mathematics 109 (2001) 253–278

DISCRETE APPLIED MATHEMATICS

On the combinatorics of leftist trees

P. Nogueira [1]

CFIF, Edifício Ciência (Física), Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 Lisboa, Portugal

Received 2 October 1996; revised 2 May 2000; accepted 30 May 2000

Abstract

Let ℓ_n be the number of leftist trees with n nodes. The corresponding (ordinary) generating function $\ell(x)$ is shown to satisfy an explicit functional equation, from which a specific recurrence for the ℓ_n is obtained. Some basic analytic properties of $\ell(x)$ are uncovered. Then the problem of determining average quantities (*expected additive weights*, in the notation of Kemp (Acta Inform. 26 (1989) 711–740)) related to the distribution of nodes is re-analysed. Finally, the average height of leftist trees is shown to be asymptotic to $n^{1/2}$, apart from a multiplicative constant that can be evaluated with high accuracy. © 2001 Elsevier Science B.V. All rights reserved.

MSC: 05C05; 05C30

1. Introduction

A tree is a connected graph without cycles, and a rooted tree is a tree in which a certain node (which is called the *root*) has been singled out. Let u and v be two nodes of a rooted tree T with root r; u is called a $*$-*subnode* of v if v belongs to the (simple) path from u to r, and it is said to be a *subnode* of v if in addition u and v are adjacent. A *leaf* is a node without subnodes. The subtree T_u of T is the tree that contains the node u (the root of T_u) and all the $*$-subnodes of u. Herein we will deal exclusively with *ordered* trees, also known as *plane* trees; these are trees in which the ordering of the sub-nodes is relevant.

There are at least two competing definitions of binary trees in the literature. It all hinges on whether internal nodes (that is, nodes that are not leaves) are required to have exactly two subnodes (a left subnode *and* a right subnode), or are allowed to have either one or two subnodes (a left one and/or a right one). Herein we adopt the latter definition, recalling at the same time that the trees satisfying the former definition may be seen as *extended binary trees* (we borrow this terminology from [12, p. 399]). Since

[1] Researcher of the Technical University of Lisbon.
E-mail address: paulo.nogueira@ist.utl.pt (P. Nogueira).

PII: S0166-218X(00)00250-X

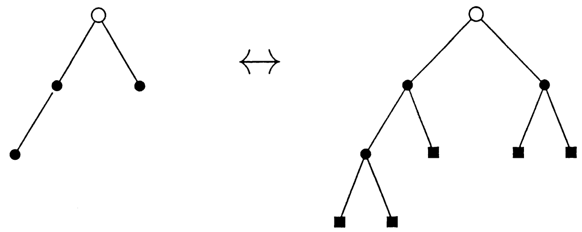

Fig. 1. An example of the correspondence between binary trees (left) and extended binary trees (right).

there is a simple bijection between both types of trees (see also Fig. 1), the results presented in this paper may be easily adjusted to match the alternative definition.

The origin of leftist trees can be traced back to the work of Crane [2], who used linked binary trees to represent priority queues in an efficient way (priority queues play an important role in various algorithms). This method was slightly modified by Knuth, who then defined leftist trees [13]. Due to the introduction of auxiliary nodes (the empty Λ nodes), that definition adjust itself to both definitions of binary trees. However, the Λ nodes are not real nodes (from a programmer's perspective, real nodes contain two fields for pointers to possible subnodes, and a Λ node simply serves to indicate the absence of a particular subnode). The Λ nodes were also ignored by Knuth when he asked for the number of leftist trees with n nodes (see [13, Section 5.2.3, exercise 34]). This evidence may serve to justify our definitions.

To define leftist trees without resorting to auxiliary nodes one may proceed as shown next. The *d-number* [2] $d(u)$ of a node u (from a binary tree), is defined recursively as follows. If u has less than 2 subnodes, then $d(u)=1$; else, $d(u)=\min(d(u_l),d(u_r))+1$, where u_l and u_r are (respectively) the left and right subnodes of u. A leftist tree is a binary tree such that for every node u with two subnodes the relation $d(u_l)\geqslant d(u_r)$ is satisfied. If the root of a leftist tree T has d-number k then we also say that the d-number of T is k, and write $d(T)=k$. Fig. 2 contains the set of leftist trees with less than 5 nodes.

By now various properties of leftist trees are already known [8,9] (there is also a very brief review in [15]). In this paper we show that the problem is not as intractable as previously thought (from a non-numerical point of view), and that the various identities satisfied by the generating functions involved can be exploited efficiently. Specifically, we find an explicit functional equation for the ordinary generating function enumerating leftist trees with n nodes and a recurrence for the coefficients of that generating function, refine some of the previously obtained (numerical) results, and obtain a few more (new) results (including the average height of leftist trees).

[2] This is equivalent to the `DIST` field defined in [13].

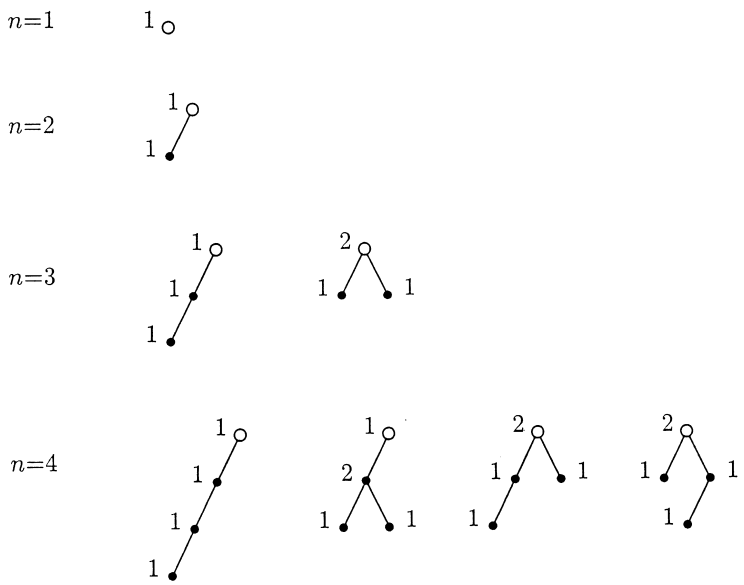

Fig. 2. The set of leftist trees with less than 5 nodes. The d-number of a node is indicated by the nearby integer.

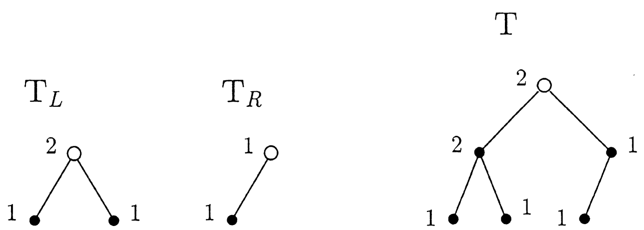

Fig. 3. A leftist tree T together with its left and right subtrees (T_L and T_R, respectively).

2. The functional equation for $\ell(x)$

Let the symbols ℓ_n and $\ell(x)$ denote the number of leftist trees with n nodes and the corresponding ordinary generating function, respectively; furthermore, let $\mathscr{L}_n$ denote the set of leftist trees with n nodes. If the leftist trees are required to have a specific d-number (k, say) then the symbols $\ell_{k,n}$, $\ell_k(x)$, and $\mathscr{L}_{k,n}$ will be used instead. Let $T_\emptyset$ be the empty leftist tree, for which we define $d(T_\emptyset)=0$; $T_\emptyset$ is the only leftist tree with null d-number, which means that $\ell_0(x)=1$. The inclusion of $T_\emptyset$ is somewhat artificial but it will lead to the simplification of several formulae.

As it happens with other kinds of trees, leftist trees can be constructed from smaller trees by a recursive scheme. A specific example is shown in Fig. 3: given two leftist trees T_L and T_R (left) such that $d(T_L)\geqslant d(T_R)$, one may obtain a larger leftist tree T

Table 1
The numbers $\ell_{k,n}$ for $n \leqslant 11$

	n										
k	1	2	3	4	5	6	7	8	9	10	11
1	1	1	1	2	4	8	17	38	87	203	482
2			1	2	4	9	20	45	104	245	585
3							1	4	12	34	93

(right) by adding one extra node (the root of T), and transforming the roots of T_L and T_R into (respectively) the left and right subnodes of the root of T; in principle T_R could be the empty tree, in which case the root of T would have a single subnode. Within our terminology T_L and T_R are called (respectively) the left and right subtrees of T. The d-number of T is $1 + d(T_R)$. A simple generalization of the previous reasoning proves the following.

Proposition 1. *The numbers $\ell_{k,n}$ may be computed recursively by means of the following relations*:

$$\ell_{k,0} = \delta_{k,0}, \quad k \geqslant 0; \quad \ell_{0,n} = \delta_{0,n}, \quad n \geqslant 0, \tag{1}$$

$$\ell_{k,n} = \sum_{j \geqslant k-1} \sum_{m=0}^{n-1} \ell_{j,m} \ell_{k-1,n-1-m}, \quad k,n \geqslant 1.$$

In the above formula the index j is effectively bounded from above by a function of n, since a leftist tree with d-number j contains at least $2^j - 1$ nodes. The numbers $\ell_{k,n}$ obtained from the previous recurrence are shown in Table 1 for $1 \leqslant n \leqslant 11$ (null values have been omitted). In principle, it is possible to compute any ℓ_n from this recurrence, since $\ell_n = \ell_{1,n+1}$. However, this is not suited for deriving asymptotic results (moreover, stopping at this point would certainly not lead to the discovery of the results displayed in the remaining of this paper).

As part of the strategy for constructing a functional equation for $\ell(x)$ we will consider first the generating functions $\ell_k(x)$. Since $\ell(x)$ can be expressed as

$$\ell(x) = \sum_{k=0}^{\infty} \ell_k(x), \tag{2}$$

the recurrence relations (1) may be condensed into

$$\ell_k(x) = x\ell_{k-1}(x) \sum_{j \geqslant k-1} \ell_j(x), \quad k \geqslant 1. \tag{3}$$

In particular, for $k = 1$, this gives

$$\ell_1(x) = x\ell(x). \tag{4}$$

A variant of (3) that may be obtained quite straightforwardly is

$$\ell_k(x) = \ell_{k-1}(x)\left(\ell_1(x) - x\sum_{j=0}^{k-2}\ell_j(x)\right), \quad k \geqslant 1. \tag{5}$$

This latter form provides a recursive scheme for the $\ell_k(x)$, that is, it may be used to express $\ell_k(x)$ (for any $k \geqslant 2$) in terms of a polynomial in x and $\ell_1(x)$. However, (5) is not completely equivalent to (3), the difference being that the constraint (4) is missing. It turns out that it is possible to solve the (infinite order) system of functional equations given by (3). The underlying recurrence (1) ensures that there is a single analytic solution. The functional equation (with constraints)

$$\ell_1(x) = x + \ell_1(x\ell_1(x)), \quad \ell_1(0) = 0 \tag{6}$$

is a contribution from out personal oracle. Before proving the correctness of this prediction (which will be done in Theorem 1) it is advantageous to find some of its implications.

Proposition 2. *If* $\ell_1(x)$ *satisfies* (6) *then*

$$\ell_1(x\ell_k(x)) = \ell_1(x) - x\sum_{j=0}^{k-1}\ell_j(x), \quad k \geqslant 0, \tag{7}$$

$$\ell_k(x) = \prod_{j=0}^{k-1}\ell_1(x\ell_j(x)), \quad k \geqslant 0, \tag{8}$$

and

$$\ell_j(x)\ell_{k-j}(x\ell_j(x)) = \ell_k(x), \quad 0 \leqslant j \leqslant k, \tag{9}$$

should hold identically.

Proof. Herein we prove only the first of these identities; the other two proofs are also simple, and the reader will surely find them. A simple inductive argument is all that is needed: if (7) holds for all integers k less than i (in particular, for $k = i-1$) then the derivation

$$\begin{aligned}\ell_1(x) - x\sum_{j=0}^{i-1}\ell_j(x) &\overset{(7)}{=} \ell_1(x\ell_{i-1}(x)) - x\ell_{i-1}(x) \\ &\overset{(6)}{=} \ell_1(x\ell_{i-1}(x)\ell_1(x\ell_{i-1}(x))) \overset{(7,5)}{=} \ell_1(x\ell_i(x))\end{aligned}$$

shows that it also holds for i itself; moreover, (7) is trivial for $k = 0$. $\square$

We are now ready to prove the main result. The functional equation for $\ell(x)$ is actually rather surprising, since it involves the unusual nested construction $\ell(\ldots\ell(x))$.

Theorem 1. *The generating function $\ell(x)$ for leftist trees with n nodes satisfies the functional equation*

$$\ell(x) = 1 + x\ell(x)\ell(x^2\ell(x)). \tag{10}$$

Proof. What has been shown up to now is this: if a sequence of generating functions $\ell_k(x)$ obeys to (5) for $k \geqslant 2$, and if $\ell_1(x)$ satisfies (6), then a number of properties follow, including (7)–(9). It would be possible to construct a different sequence by choosing a distinct $\ell_1(x)$ and then apply the recurrence (5). However, (4) cannot be fulfilled unless $\ell_1(x)$ is the counting series for leftist trees with d-number equal to 1. The derivation

$$\begin{aligned} x\ell(x) &\overset{(2)}{=} x + x\sum_{k=1}^{\infty} \ell_k(x) \overset{(9)}{=} x + x\sum_{k=1}^{\infty} \ell_1(x)\ell_{k-1}(x\ell_1(x)) \\ &\overset{(2)}{=} x + x\ell_1(x)\ell(x\ell_1(x)) \end{aligned}$$

implies that a solution of the functional equation

$$f(x) = x + f(x\ell_1(x)) \tag{11}$$

is given by $f(x)=x\ell(x)$. The next step is to realize that there is a unique formal power series $f(x) = \sum_{n=1}^{\infty} f_n x^n$ (the constant term has been deliberately excluded) that is a solution of (11). The proof relies on the following observation: if f is replaced by its power series representation then Eq. (11), in that exact form, leads to a recurrence relation for the coefficients f_n (for $n > 1$) and simultaneously determines f_1.

We may now conclude that (4) is fulfilled by assuming (6). Therefore, due to the uniqueness of the solution, we may assert the validity of (6), from which the theorem follows easily. □

It is now possible to obtain a recurrence relation for the coefficients ℓ_n which does not involve the $\ell_{k,n}$.

Corollary 1. *The numbers ℓ_n follow the recurrence relation*

$$\ell_{m-1} = \sum_{(j)} \delta_{0,j_1} \frac{(j_2+j_3+\cdots+j_m)!}{j_2!j_3!\cdots j_m!} \ell_0^{j_2}\ell_1^{j_3}\ell_2^{j_4}\cdots\ell_{m-2}^{j_m}\ell_{j_2+j_3+\cdots+j_m-1} \tag{12}$$

for all $m \geqslant 2$, with $\ell_0 \equiv 1$. The summation extends over the partitions $(1^{j_1}2^{j_2}\ldots m^{j_m})$ of m, with j_k being the number of parts equal to k.

Proof. This may be achieved by taking an arbitrary derivative of (10) with respect to x (actually differentiating (6) is simpler). Here one may find helpful the general result

$$\frac{1}{m!}\frac{\mathrm{d}^m}{\mathrm{d}x^m} f(g(x)) = \sum_{(j)} \left(f^{(s)}(g(x)) \prod_{k=1}^{m} \frac{(g^{(k)}(x))^{j_k}}{j_k!k!^{j_k}} \right) \tag{13}$$

(with $s = j_1 + j_2 + \cdots + j_m$), which is often referred to as Faà di Bruno's formula (see e.g. [12]). □

3. The functions $\ell(x)$ and $\ell_k(x)$

It is known that the ordinary generating function for the number of binary trees with n nodes is analytic at $x=0$, and that its radius of convergence is equal to 1/4. The fact that every leftist tree is a binary tree already implies that $\ell(x)$ must be analytic at $x=0$, and that its radius of convergence ρ is at least 1/4; in addition, the existence of leftist trees with an arbitrarily large number of nodes indicates that $\rho \leqslant 1$. Other properties (mostly analytical ones) of $\ell(x)$ and its variants are summarized in this section.

Proposition 3. *The power series $\ell_1(x)$ converges, for $x=\rho$, and $\ell_1(\rho) < 1$.*

Proof. The function $\ell_1(x)$ is clearly continuous and strictly increasing in the interval $[0, \rho)$. Setting $\ell_1(x)=1$ in (6) gives $x=0$, which is a contradiction. Hence $\lim_{x\to\rho^-} \ell_1(x)$ exists and is bounded from above by 1. Now one of the Tauberian theorems[3] gives

$$\ell_1(\rho) = \sum_{n=1}^{\infty} \ell_{1,n}\rho^n = \lim_{x \to p^-} \ell_1(x), \tag{14}$$

and the left continuity of $\ell_1(x)$ for $x=\rho$ suffices to exclude the case $\ell(\rho)=1$. $\square$

Proposition 4. *The inequality*

$$\ell_k(x_1) < \left(\frac{x_1}{x_2}\right)^{2^k-1} \ell_k(x_2) \tag{15}$$

holds for all $k \geqslant 1$ and $x_1, x_2 \in (0, \rho]$ such that $x_1 < x_2$.

Proof. The leading term in $\ell_1(x)$ is equal to x, and in general the leading term in $\ell_k(x)$ equals x^{2^k-1}, by virtue of (3). This means that $\ell_k(x)/x^{2^k-1}$, being a power series with positive coefficients only, represents a strictly increasing function in $[0, \rho]$ – and now (15) follows at once. $\square$

Proposition 5. *Let $\ell_1^{-1}(x)$ denote the inverse of $\ell_1(x)$, that is, the generating function for which $\ell_1(\ell_1^{-1}(x)) = x$ holds identically; then*

$$\ell_1^{-1}(x) = -\ell_1(-x). \tag{16}$$

Proof. The functional equation

$$f(z) = z - \ell_1(zf(z)) \tag{17}$$

has at most a single solution $f(z)$ (analytic in a neighbourhood of zero, that is). The proof is quite similar to the one given for the functional equation (11) and because of this details will be omitted.

[3] Specifically: if (a) $\lim_{x\to 1^-} \sum_{n=0}^{\infty} a_n x^n$ exists and equals A, and (b) $a_n \geqslant 0$ for all n, then $\sum_{n=0}^{\infty} a_n$ converges and has sum A.

It is well known that if a function ϕ is analytic at $w_0 \in \mathbf{C}$ and $\phi'(w_0) \neq 0$ then ϕ has a local inverse which is single-valued and analytic. This implies that there is a disk $|z| < \sigma$ where the relation $z = \ell_1(w)$ can be inverted to $w = \ell_1^{-1}(z)$. Now let $f(z) = \ell_1^{-1}(z)$ and rewrite (17) in terms of w; the result is just a rewriting of (6), which shows that $\ell_1^{-1}(z)$ is a solution of (17).

As it happens, a second solution may be displayed. If one makes the substitution $f(z) = -\ell_1(-z)$ in (17) one will simply find (6) with x replaced by $-z$. Hence $-\ell_1(-z)$ is also a solution of (17). The proof is now complete, due to the uniqueness of the solution. □

It should be noted that the property stated in Proposition 5 is satisfied by an infinity of functions, for example by any function of the form $F_a(x) = x/(1-ax)$, with a being a constant. Translating (16) in terms of $\ell(x)$ and $\ell_1'(x)$ yields

$$\ell(x)\ell(-\ell_1(x)) = 1, \quad \ell_1'(x)\ell_1'(-\ell_1(x)) = 1. \tag{18}$$

Proposition 6. *The radius of convergence ρ of the power series $\ell_1(x)$ satisfies*

$$\rho\ell_1'(\rho\ell_1(\rho)) = 1. \tag{19}$$

Proof. Analyticity breaks down on the circle of convergence at those points where the derivative does not exist. Taking the derivative of (6) with respect to x gives

$$\ell_1'(x) = \frac{1 + \ell_1(x)\ell_1'(x\ell_1(x))}{1 - x\ell_1'(x\ell_1(x))}. \tag{20}$$

A singularity will occur if the denominator of (20) vanishes and the numerator does not. Due to the positivity of the coefficients $\ell_{1,n}$ one has that $\rho\ell_1'(\rho\ell_1(\rho)) > |z\ell_1'(z\ell_1(z))|$ for any $z \neq \rho$ on the circle of convergence, and thus (19) is satisfied on the circle of convergence for $z = \rho$ only (note that $z\ell_1(z)$ is inside the circle of convergence if $|z| \leqslant \rho$). □

Proposition 7. *Let $g(y) = \ell_1^{-1}(y)$ for $0 \leqslant y \leqslant \ell_1(\rho)$. Then*

$$\eta_1 \equiv \lim_{y \to \ell_1(\rho)^-} g'(y) = 0, \qquad \eta_2 \equiv \lim_{y \to \ell_1(\rho)^-} -g''(y) = \frac{\rho^3 \ell_1''(\rho\ell_1(\rho))}{\rho + \ell_1(\rho)}. \tag{21}$$

Proof. An expression for $\ell_1''(x)$ may be obtained from (20). After differentiating with respect to x and making some rearrangements one may find that

$$\ell_1''(x) = \frac{[x + \ell_1(x)]^2}{[1 - x\ell_1'(x\ell_1(x))]^3}\ell_1''(x\ell_1(x)) + \frac{2\ell_1'(x)\ell_1'(x\ell_1(x))}{1 - x\ell_1'(x\ell_1(x))}. \tag{22}$$

It is well known that the derivatives of $g(y)$ evaluated at $y = y_0 = \ell_1(x_0)$ can be expressed in terms of the derivatives of $\ell_1(x)$ evaluated at $x = x_0$. Combining those classical results with (19), (20), and (22) yields (21). □

Whether ρ, $\ell_1(\rho)$, etc., may be expressed by means of "simple" formula appears to be a hard problem. The numerical approximation of these quantities seems unavoidable, and a method for doing that will be described in Appendix A. The results are summarized below.

Proposition 8. *The numerical values of* ρ, $\ell_1(\rho)$, $\rho\ell_1(\rho)$, $\ell_1(\rho\ell_1(\rho))$, $\ell_1'(\rho\ell_1(\rho))$, *and* $\ell_1''(\rho\ell_1(\rho))$ are given by

$$\begin{aligned}
&\rho = 0.36370409\ldots,\\
&\ell_1(\rho) = 0.82329973\ldots,\\
&\rho\ell_1(\rho) = 0.29943748\ldots,\\
&\ell_1(\rho\ell_1(\rho)) = 0.45959564\ldots,\\
&\ell_1'(\rho\ell_1(\rho)) = 2.7494879\ldots,\\
&\ell_1''(\rho\ell_1(\rho)) = 22.784216\ldots\ .
\end{aligned} \tag{23}$$

4. Additive weights and systems of functional equations

The general analysis of *additive weights* (in the sense of [9]) requires one to study the system of functional equations

$$\begin{aligned}
&\mu_0(x) = \beta_0(x),\\
&\mu_k(x) = \beta_k(x) + \gamma_1 x\ell_{k-1}(x)\left(\mu(x) - \sum_{j=0}^{k-2}\mu_j(x)\right)\\
&\qquad + \gamma_2 x\mu_{k-1}(x)\left(\ell(x) - \sum_{j=0}^{k-2}\ell_j(x)\right), \quad k > 0,
\end{aligned} \tag{24}$$

where

$$\mu(x) = \sum_{k=0}^{\infty}\mu_k(x). \tag{25}$$

The coefficient $[x^n]\mu_k(x)$ is equal to the sum of the weights of all leftist trees with n nodes and d-number k. Any particular instance of this system is specified by two constants (γ_1 and γ_2) and an infinite sequence of generating functions $\beta_k(x)$ (in practice $\beta_0(x)$ is a constant, not an arbitrary generating function). A simple inductive reasoning will show that the relation

$$\mu_k(x) = a_k(x)\mu(x) + \sum_{j=0}^{k} b_{k,j}(x)\beta_j(x), \quad k \geqslant 0 \tag{26}$$

must hold for appropriate formal power series $a_k(x)$ and $b_{k,j}(x)$ (which may be expressed in terms of γ_1, γ_2, x, and the $\ell_i(x)$). In order to simplify the notation we assume that γ_1 and γ_2 have been defined *a priori*, and therefore omit the explicit dependence on those constants. One may actually find several properties of the $a_k(x)$ and $b_{k,j}(x)$;

for example, substituting (26) into (24) leads to the recurrences

$$a_k(x) = \gamma_1 x\ell_{k-1}(x)\left(1 - \sum_{j=0}^{k-2} a_j(x)\right) + \gamma_2 x a_{k-1}(x)\left(\ell(x) - \sum_{j=0}^{k-2} \ell_j(x)\right), \quad k > 0 \tag{27}$$

with $a_0(x) = 0$, and

$$b_{i,j}(x) = \delta_{i,j} - \gamma_1 x\ell_{i-1}(x)\sum_{k=0}^{i-2} b_{k,j}(x) + \gamma_2 \ell_1(x\ell_{i-1}(x)) b_{i-1,j}(x), \quad i \geqslant 1,\ j \geqslant 0, \tag{28}$$

with $b_{0,0}(x) = 1$ (note that $b_{i,j}(x) = 0$ if $i < j$). Using induction, one can easily show that the latter recurrence implies

$$b_{i,j}(x) = b_{i-1,j-1}(x\ell_1(x)), \quad i,j > 0. \tag{29}$$

By adding up (26) over k we can obtain a formal solution for $\mu(x)$, namely

$$\mu(x) = (1 - f_0(x))^{-1} \sum_{j=0}^{\infty} g_j(x)\beta_j(x), \tag{30}$$

with

$$f_0(x) = \sum_{k=1}^{\infty} a_k(x), \qquad g_j(x) = \sum_{k=0}^{\infty} b_{k,j}(x). \tag{31}$$

Of course, we have to show that these are proper definitions, in the sense that the above series give rise to well-defined formal power series. From the analysis of the recurrences (27) and (28) it is obvious that the leading terms of $a_k(x)$ and $b_{k,j}(x)$ are proportional to $x^{2^{k-1}}$ (or to a higher power of x) for $k \geqslant 1$ and $k > j$, respectively. This is more than enough for the convergence of (31) in the ring of formal power series. Using standard methods it can be shown that $f_0(x)$ and the $g_k(x)$ (for $k \geqslant 0$) are analytic functions, at least for $x \in I = [0, \rho)$.

It turns out that one can derive a few explicit closed form relations for the functions $f_0(x)$ and $g_k(x)$, still for arbitrary γ_1 and γ_2. The simplest one is a recurrence that follows trivially from (31) and (29):

$$g_{k+1}(x) = g_k(x\ell_1(x)). \tag{32}$$

Additional relations may be obtained by substituting explicit solutions of (24) into (30), and then combining the resulting expressions. Coincidentally, there are choices of the generating functions $\beta_k(x)$ such that both $\beta_k(x)$ and $\mu_k(x)$ vanish but for a finite number of values of k, e.g.

$$\begin{aligned} \beta_0(x) &= \mu(x), \\ \beta_1(x) &= -(\gamma_1 x + \gamma_2 \ell_1(x))\mu(x), \\ \mu_0(x) &= \mu(x), \end{aligned} \tag{33}$$

and

$$\begin{aligned}\beta_1(x) &= (1-\gamma_1 x)\mu(x),\\ \beta_2(x) &= -(\gamma_1 x\ell_1(x) + \gamma_2\ell_1(x\ell_1(x)))\mu(x),\\ \mu_1(x) &= \mu(x)\end{aligned} \tag{34}$$

(in each case only the non-zero $\beta_k(x)$ and $\mu_k(x)$ are shown). The fact that these are always solutions of (24) implies the following relations:

$$1 - f_0(x) = g_0(x) - (\gamma_1 x + \gamma_2\ell_1(x))g_0(x\ell_1(x)), \tag{35}$$

$$g_0(x) = (1+\gamma_2\ell_1(x))g_0(x\ell_1(x)) - (\gamma_1 x\ell_1(x) + \gamma_2\ell_1(x\ell_1(x))g_0(x\ell_2(x)). \tag{36}$$

Further progress seems to be hard, unless one starts to deal with specific cases; still, numerical computing is often needed. Nevertheless – this will be shown in next section – in some cases one can really find simple, neat formulae.

5. Expected additive weights: examples

The expected additive weights presented in this section can be determined using the iterative numerical procedure presented in [9] (some of the numerical values have actually been given in that paper). Nevertheless, we will show that here and there one can gain further insight by using generating functions as much as possible, not to mention that not all the results given here are expected additive weights. Thus, new results are interspersed with known ones, but the latter ones cannot be omitted altogether for the benefit of a coherent presentation.

To extract the asymptotic behaviour of the coefficients of a generating function we use the standard method, which had its origin in Pólya's work [17,18]. An improvement was made by Otter [16], and subsequent expositions include [1,6,7].

To determine (asymptotic) expected additive weights of leftist trees one needs the asymptotic number of leftist trees (with n nodes). Although this is a known result we restate it here in a slightly different form (the multiplicative constant can be written in terms of constants related to $\ell_1(x)$ and its second derivative).

Proposition 9. *As n tends to infinity, the number of leftist trees in $\mathscr{L}_n$ is asymptotic to*

$$\frac{c_1 b^n}{n^{3/2}}. \tag{37}$$

Here $b = \rho^{-1} = 2.7494879\ldots$, *and* $c_1 = (2\pi\rho\eta_2)^{-1/2} = 0.68837122...$ *with* η_2 *being defined in* (21).

Proof. It should be clear that $\ell_1^{-1}(y)$ can be expanded in a left neighbourhood of $y = \ell_1(\rho)$ (for details see the references given above). Starting from (21) one may find

$$\ell(x) = \ell(\rho) - \left(\frac{2}{\rho^2 \eta_2}\right)^{1/2} (\rho - x)^{1/2} + \cdots, \tag{38}$$

and then a simple application of Darboux's theorem (see e.g. [15]) proves (37). □

This may be considered a typical asymptotic result, at least if one looks at the results obtained for other types of rooted trees (see, for instance, [17,16,5,7]). Proposition 9 is in agreement with [8], since the apparent discrepancy in the constant factor is simply due to a different definition of the trees being counted. Here, as in the original formulation of the problem, the parameter n denotes the number of (internal) nodes while in [8] it denotes the number of leaves in the extended tree. These differ by 1, and that is why the constant α in [8] is equal to $a\rho$. The relation between Kemp's generating functions and the ones used in this paper is $H(z) = z\ell(z)$ and $T_k(z) = z\ell_{k-1}(z)$.

To simplify the definition of the generating functions introduced hereafter we adopt the following rule. Whenever we associate a sequence – F_n, say – to some (not necessarily additive) weight in $\mathscr{L}_n$, then it will be implicit that the double indexed quantities $F_{k,n}$ enumerate that weight in $\mathscr{L}_{k,n}$. Furthermore, it will be understood that the corresponding generating functions are $F(x)$ and $F_k(x)$, that is,

$$F(x) = \sum_{n=1}^{\infty} F_n\, x^n, \qquad F_k(x) = \sum_{n=1}^{\infty} F_{k,n}\, x^n. \tag{39}$$

Alternatively, once we define $F(x)$ none of the other quantities (that is, F_n, $F_k(x)$, and $F_{k,n}$) will need to be declared explicitly.

5.1. The average fraction of leftist trees with d-number equal to k

Let us define an infinite sequence $\{\alpha_k\}$ as follows:

$$\alpha_0 = 0,$$

$$\alpha_k = \rho\alpha_{k-1}\left(\ell(\rho) - \sum_{j=0}^{k-2} \ell_j(\rho)\right) + \rho\ell_{k-1}(\rho)\left(1 - \sum_{j=0}^{k-2} \alpha_j\right), \quad k > 0. \tag{40}$$

The next proposition presents some properties of the terms of that sequence. Results similar to (41) and (42) may be found in Ref. [9] (in the notation of that paper, the term α_k may be written as $h_{k+1}(\eta, a)$).

Proposition 10. *The numbers α_k have the following properties*:

$$\alpha_k = \lim_{x \to \rho^-} \frac{\ell'_k(x)}{\ell'(x)}, \quad k \geqslant 0, \tag{41}$$

$$\sum_{k=1}^{\infty} \alpha_k = 1, \tag{42}$$

$$\alpha_k = \rho\ell_{k-1}(\rho\ell_1(\rho)) + \rho^2\ell_1(\rho)\ell'_{k-1}(\rho\ell_1(\rho)), \quad k \geqslant 1, \tag{43}$$

$$\alpha_k = (\ell_1(\rho\ell_{k-1}(\rho)) + \rho\ell_{k-1}(\rho)\ell'_1(\rho\ell_{k-1}(\rho)))\alpha_{k-1}, \quad k \geqslant 2. \tag{44}$$

Proof. In order to prove (41) one may start by taking the derivative of (3) with respect to x and dividing the result by $\ell'(x)$. Then one will realize that the recurrence for the numbers $\lim_{x\to\rho^-}\ell'_k(x)/\ell'(x)$ is identical to (40) (note also that $\ell'_0(\rho)=0$). As for (42), it follows trivially from (41).

The proof of (43) goes like this: setting $j = 1$ in identity (9) gives an expression for $\ell_k(x)$ which allows a direct evaluation of the limit in (41). Eq. (44) can be proved similarly (this time with $j = k - 1$). □

It is now clear that, for fixed k, and in agreement with [9],

$$\ell_{k,n} \sim \alpha_k\ell_n. \tag{45}$$

However, one may add a bit more about the infinite sequences $\{\ell_k(\rho)\}$ and $\{\alpha_k\}$. Evidently they both converge to zero, since they are sequences of positive terms (at least for $k > 0$) whose series converge. This is, however, a weak statement, since one can easily prove – using (9) and (44) – that

$$\lim_{k\to\infty}\frac{\ell_{k+1}(\rho)}{\ell_k(\rho)} = \lim_{k\to\infty}\frac{\alpha_{k+1}}{\alpha_k} = 0. \tag{46}$$

In fact, it is possible to find the asymptotic behaviour of both $\ell_k(\rho)$ and α_k as $k \to \infty$ (as it is known, α_k equals the probability that a random leftist tree with n nodes has d-number equal to k, in the limit $n \to \infty$).

Theorem 2. *The asymptotic expansions*

$$\ell_k(\rho) \sim \rho^{-1}\omega^{2^k}(1 - \tfrac{1}{2}\omega^{2^k} + \mathrm{O}((\omega^{2^k})^2)) \tag{47}$$

and

$$\alpha_k \sim c_\alpha 2^k\omega^{2^k}(1 - \omega^{2^k} + \mathrm{O}((\omega^{2^k})^2)) \tag{48}$$

hold as $k \to \infty$, for some positive constants ω and c_α.

Proof. We give only an outline of the proof, but the details omitted are elementary. After noticing that

$$[\rho\ell_k(\rho)]^{1/2^k} = \rho\prod_{i=0}^{k-1}\left(\frac{\ell_1(\rho\ell_i(\rho))}{\rho\ell_i(\rho)}\right)^{1/2^{i+1}} \tag{49}$$

holds, one may observe that the infinite product appearing in this last equation converges (one easy way to prove this involves the use of inequality (15), for $k = 1$). Hence we may define

$$\omega = \rho\prod_{i=0}^{\infty}\left(\frac{\ell_1(\rho\ell_i(\rho))}{\rho\ell_i(\rho)}\right)^{1/2^{i+1}}, \tag{50}$$

Table 2
Numerical data

k	$\ell_k(\rho)$	α_k	λ_k
0	1	0	
1	0.82329973...	0.36370409...	1
2	0.37838497...	0.46659430...	0.64175126...
3	$0.60601442\ldots\times 10^{-1}$	0.16229603...	0.53955198...
4	$0.13658352\ldots\times 10^{-2}$	$0.73982197\ldots\times 10^{-2}$	0.50557999...
5	$0.67882937\ldots\times 10^{-6}$	$0.73557580\ldots\times 10^{-5}$	0.50012422...
6	$0.16759827\ldots\times 10^{-12}$	$0.36321719\ldots\times 10^{-11}$	0.50000006...
7	$0.10216150\ldots\times 10^{-25}$	$0.44280664\ldots\times 10^{-24}$	0.50000000...

and now a simple analysis shows that

$$[\rho\ell_k(\rho)]^{1/2^k} = \omega + \mathrm{O}\left(\frac{\ell_k(\rho)}{2^k}\right). \tag{51}$$

This shows the correctness of the leading term in (47). The result for the α_k may be obtained rather similarly. The first point is that one may write

$$\frac{\alpha_k}{2^k\rho\ell_k(\rho)} = \frac{1}{2\ell_1(\rho)}\prod_{i=1}^{k-1}\left(\frac{\ell_1(\rho\ell_i(\rho)) + \rho\ell_i(\rho)\ell_1'(\rho\ell_i(\rho))}{2\ell_1(\rho\ell_i(\rho))}\right)^{1/2^{i+1}} \tag{52}$$

and the second one is that

$$c_\alpha = \frac{1}{2\ell_1(\rho)}\prod_{i=1}^{\infty}\left(\frac{\ell_1(\rho\ell_i(\rho)) + \rho\ell_i(\rho)\ell_1'(\rho\ell_i(\rho))}{2\ell_1(\rho\ell_i(\rho))}\right)^{1/2^{i+1}} \tag{53}$$

is a proper definition (i.e., the infinite product converges). Non-leading terms of those asymptotic expansions may be determined iteratively by estimating the finite products to increasingly higher order. □

Numerically, $\omega = 0.62160700\ldots$ and $c_\alpha = 0.93104077\ldots$. However, the computation was based on a number of initial terms of the sequences $\{\ell_k(\rho)\}$ and $\{\alpha_k\}$; for that reason, and also because numerical errors tend to accumulate rather fast, the use of those two constants for predicting (with high accuracy) further terms of the above mentioned sequences is somewhat limited.[4] Nonetheless, the previous theorem shows that the infinite sequences $\{\ell_k(\rho)\}$ and $\{\alpha_k\}$ approach zero very fast (see also Table 2), and also that the relative error associated with the leading term of the asymptotic expansions (47) and (48) decreases rapidly with k. As it will be seen later, some expected additive weights can be expressed by formulae containing some kind of infinite series (or product) whose general term depends on those sequences; while a general statement might be out of reach, in the cases we looked at the formulae could be evaluated swiftly and accurately.

[4] There are a few hints pointing to the existence of a power series (i.e., function) $H(x)$ such that $H(x)\ell_1(H(x)) = H(x^2), \rho\ell_k(\rho) = H(\omega^{2^k})$, and $H'(\omega) = 0$. This might lead to an alternative (even if computationally less efficient) determination of ω.

5.2. The average fraction of left and right subnodes

If a node u (in some leftist tree) has a right subnode then u also has a left subnode. The converse is not true though, which means that the left subnodes of a leftist tree will often outnumber the right subnodes. Let $L(x)$ and $R(x)$ enumerate (respectively) the left and right subnodes in *all* the leftist trees with n nodes. The total number of nodes (which may be divided into left subnodes, right subnodes, and roots) in $\mathscr{L}_n$ is equal to $n\,\ell_n$, and thus

$$L(x) + R(x) + \ell(x) - 1 = x\,\ell'(x). \tag{54}$$

Similarly,

$$L_k(x) + R_k(x) + \ell_k(x) = x\,\ell_k'(x) \tag{55}$$

must hold for every $k \geqslant 1$. Analogously to the case of the coefficients $\ell_{k,n}$, there exist recurrence relations for the coefficients $L_{k,n}$ and $R_{k,n}$. The recursive construction of leftist trees (see Fig. 3) leads to

$$\begin{aligned} &R_{k,1} = 0, \quad k \geqslant 1, \\ &R_{1,n} = \sum_k R_{k,n-1}, \quad n > 1, \\ &R_{k,n} = \ell_{k,n} + \sum_{j \geqslant k-1} \sum_{m=1}^{n-2} (\ell_{j,m} R_{k-1,n-1-m} + R_{j,m} \ell_{k-1,n-1-m}), \quad k,n > 1. \end{aligned} \tag{56}$$

It follows from here that the equations for $R(x)$ and the $R_k(x)$ are a particular instance of the ansatz (24), in which case one has

$$\gamma_1 = \gamma_2 = 1, \quad \beta_k(x) = \begin{cases} 0 & (k = 0, 1) \\ \ell_k(x) & (k \geqslant 2) \end{cases} \quad \Rightarrow \quad \mu_k(x) = R_k(x). \tag{57}$$

Performing a similar analysis for $L(x)$ gives

$$\gamma_1 = \gamma_2 = 1, \quad \beta_k(x) = \begin{cases} 0 & (k = 0) \\ \ell_1(x) - x & (k = 1) \\ \ell_k(x) & (k \geqslant 2) \end{cases} \quad \Rightarrow \quad \mu_k(x) = L_k(x). \tag{58}$$

The equations for $\ell(x)$ and the $\ell_k(x)$ may also be written in that form, even if that is somewhat unnatural:

$$\gamma_1 = \gamma_2 = 1, \quad \beta_k(x) = \begin{cases} 1 & (k = 0) \\ \\ -\ell_k(x) & (k \geqslant 1) \end{cases} \quad \Rightarrow \quad \mu_k(x) = \ell_k(x). \tag{59}$$

At this point we may combine the previous (somewhat faint) results into something more substantial. By blending (30), (35), (54), and (57)–(59) one may (a) obtain exact expressions for $R(x)$ and $L(x)$ in terms of $\ell(x)$ only (therefore eliminating the dependence on $f_0(x)$ and the $g_k(x)$), and (b) show that (36) can be replaced by a two-term relation (if $\gamma_1 = \gamma_2 = 1$). The two-term relation is

$$g_0(x) = G(x)g_0(x\ell_1(x)), \quad \gamma_1 = \gamma_2 = 1, \tag{60}$$

where

$$G(x) = \frac{(\ell_1(x) + x)\ell_1'(x)}{\ell_1'(x) - 1} = \ell_1(x) + \frac{1}{\ell_1'(x\ell_1(x))}. \tag{61}$$

Also (35) can be simplified, reducing to

$$g_0(x) = (1 - f_0(x))\ell_1'(x), \quad \gamma_1 = \gamma_2 = 1. \tag{62}$$

Proposition 11. *The average fraction of left subnodes is asymptotic to*

$$\frac{\ell_1(\rho)}{\ell_1(\rho) + \rho} = 0.69359484\cdots \tag{63}$$

as n tends to infinity. This result holds even if one restricts the analysis to leftist trees with fixed d-number.

Proof. As explained above, it is possible to find very convenient expressions for $L(x)$, $R(x)$, $R_k(x)$ and $L_k(x)$. For example,

$$L(x) = \frac{x\ell(x)\ell'(x) - \ell(x) + 1}{\ell(x) + 1}, \tag{64}$$

$$L_k(x) = \frac{x(\ell_1(x)\ell_k'(x) - \ell_k(x))}{\ell_1(x) + x}. \tag{65}$$

One more application of Darboux's theorem completes the proof. □

5.3. *The average fraction of nodes with a given degree*

The problem which has just been solved asked for the number of left and right subnodes belonging to the leftist trees in either $\mathscr{L}_n$ or $\mathscr{L}_{k,n}$. But one might as well have asked for the number of nodes having a certain number of subnodes (whether those subnodes are left or right subnodes is fixed by the number of subnodes). We will now show that this problem may be solved in terms of the original one.

The generating functions counting the number of nodes having two, one, and zero subnodes will be distinguished by the symbols X, Y, and Z, respectively. The identities

$$X(x) + Y(x) + Z(x) = x\ell'(x), \qquad X_k(x) + Y_k(x) + Z_k(x) = x\ell_k'(x) \tag{66}$$

hardly need a detailed justification (see (54) and (55)). The following statements are also quite obvious, given the definition of leftist trees. If a node u has two subnodes then it has a right subnode v; vice versa, if v is a right subnode of u then u has two subnodes. This shows that

$$X_k(x) = R_k(x), \qquad X(x) = R(x). \tag{67}$$

In addition, if v is a left subnode of u then u has either one or two subnodes; conversely, if a node u has at least one subnode then it must have a left subnode v. Hence

$$X_k(x) + Y_k(x) = L_k(x), \qquad X(x) + Y(x) = L(x). \tag{68}$$

Proposition 12. *Let β_s be the limit fraction – as $n \to \infty$ – of the nodes in the trees in $\mathscr{L}_n$ having exactly s subnodes. Then*

$$\beta_0 = \beta_2 = \frac{\rho}{\ell_1(\rho) + \rho} = 0.30640515\ldots, \tag{69}$$

$$\beta_1 = \frac{\ell_1(\rho) - \rho}{\ell_1(\rho) + \rho} = 0.38718969\ldots. \tag{70}$$

These limits remain unchanged even if one restricts the analysis to leftist trees with a fixed d-number.

Proof. Exact expressions for $X(x)$, $Y(x), \ldots, Z_k(x)$, may be obtained with the help of equations (66) through (68), e.g.

$$Y_k(x) = \frac{(\ell_1(x) - x)(x\ell_k'(x) + \ell_k(x))}{\ell_1(x) + x}. \tag{71}$$

Now use Darboux's theorem once more. □

5.4. The average node depth

The depth of a node is the distance from that node to the root of the tree; depending on the way one draws trees, it might be called altitude as well [14]. The value of the average node depth in $\mathscr{L}_n$ gives an estimate of the average number of steps needed to reach a node in a leftist tree with n nodes, starting from the root (and it also gives a lower bound on the average height of such trees). Herein, as in all the examples where *averages* are evaluated, all leftist trees with n nodes are assumed to be equally likely to occur.

Let us construct the generating function $s(x)$ by

$$s_n \equiv [x^n]s(x) \equiv \sum_{T \in \mathscr{L}_n} \sum_{u \in T} d(u, r). \tag{72}$$

This means that s_n is equal to the sum of the depths of all the nodes in all the leftist trees in $\mathscr{L}_n$. The recurrence relations for the $s_{k,n}$ lead to the functional equations

$$s_k(x) = x\ell_{k-1}(x)\left(s(x) - \sum_{j=0}^{k-2} s_j(x)\right) + xs_{k-1}(x)\left(\ell(x) - \sum_{j=0}^{k-2} \ell_j(x)\right) + x\ell_k'(x) - \ell_k(x), \quad k > 0. \tag{73}$$

Unlike what has been seen so far, in this case there won't be any impressive simplifications (in a sense, this example illustrates the typical situation). The final result still involves one infinite series, but one that converges fast. At this stage we need to determine a few more properties of $g_0(x)$.

Proposition 13. *The function $g_0(x)$ is finite at $x=\rho$ and (for $\gamma_1=\gamma_2=1$)*

$$g_0(\rho)=\frac{\ell_1(\rho)+\rho}{\rho}\prod_{k=1}^{\infty}(1-\rho\ell_k(\rho)\ell_1'(\rho\ell_{k+1}(\rho)))=1.6158202\ldots. \tag{74}$$

Proof. The right-hand side of (36) (this also applies to (60)) is continuous in the limit $x\to\rho^-$, hence $g_0(\rho)$ exists. By iterating (60) (and observing that $g_0(0)=1$) one may derive that

$$g_0(\rho)=\prod_{k=0}^{\infty}G(x_k),\quad \gamma_1=\gamma_2=1, \tag{75}$$

where

$$x_n=\rho\ell_n(\rho). \tag{76}$$

Using (61) and (20) one may find

$$G(x_n)=\frac{(1-x_n\ell_1'(x_{n+1}))\ell_1'(x_n)}{\ell_1'(x_{n+1})},\quad n\geqslant 1. \tag{77}$$

Substituting this result into the infinite product yields (74). The numerical evaluation can be done easily with the help of the previously computed numerical values. □

Proposition 14. *The average depth of the nodes belonging to the trees in $\mathscr{L}_n$ is asymptotic to*

$$\xi_1\sqrt{\pi n} \tag{78}$$

as $n\to\infty$, where

$$\xi_1=\frac{1}{\rho^2 g_0(\rho)}\left(\frac{\ell_1(\rho)+\rho}{2\ell_1''(\rho\ell_1(\rho))}\right)^{1/2}\sum_{k=1}^{\infty}\alpha_k g_k(\rho)=0.90674685\cdots \tag{79}$$

and with $\gamma_1=\gamma_2=1$ being implicit in the definition of the $g_k(x)$.

Proof. We are now ready to attack (73). In this case we have

$$\beta_k(x)=\begin{cases}0 & (k=0)\\ x\ell_k'(x)-\ell_k(x) & (k\geqslant 1)\end{cases} \tag{80}$$

and thus

$$s(x)=\frac{\sum_{k=1}^{\infty}(x\ell_k'(x)-\ell_k(x))g_k(x)}{1-f_0(x)}. \tag{81}$$

The leading terms of the asymptotic expansion of $f_0(x)$ as x tends to ρ^- follows from (62), and are given by

$$f_0(x)=1-g_0(\rho)\sqrt{2\eta_2}(\rho-x)^{1/2}+\cdots, \tag{82}$$

where η_2 has been defined in (21). Finding the leading term of the asymptotic expansion of $s(x)$ is now straightforward, and then one further application of Darboux's theorem settles the proposition (which is in numerical agreement with [9]). □

It is now elementary to show that the estimate (78) is also valid in the case where leftist trees are required to have a definite (positive) d-number k. The details of the proof are left to the reader.

5.5. The average size of the left and right subtrees (at the root)

Every leftist tree may be decomposed into its left and right subtrees, and its root (see Fig. 3). One may now ask what is the average size of those subtrees, in both $\mathscr{L}_n$ and $\mathscr{L}_{k,n}$. Let $|T|$ be the size of T (that is, the number of nodes in T), and define

$$E_n \equiv [x^n]E(x) \equiv \sum_{T\in\mathscr{L}_n} |T_L|. \tag{83}$$

To count the nodes in the right subtrees we will use the symbol D instead (hence we will have D_n, $D(x)$, etc).

Proposition 15. *The generating functions $E_k(x)$ may be written in the form*

$$E_k(x) = x^2\ell_{k-1}(x)\left(\ell'(x) - \sum_{j=1}^{k-2} \ell'_j(x)\right). \tag{84}$$

Proof. The equations

$$E_{1,1} = 0,$$

$$E_{1,n} = (n-1)\ell_{1,n}, \quad n\geqslant 1,$$

$$E_{k,n} = \sum_{j\geqslant k-1}\sum_{m=1}^{n-2} m\ell_{j,m}\ell_{k-1,n-1-m}, \quad k,n>1, \tag{85}$$

can be easily proved if one recalls the recursive construction of leftist trees illustrated in Fig. 3. Now multiply by x^n and sum over n (note that in this example we have $\gamma_1=\gamma_2=0$). $\square$

Finding out the generating functions $D_k(x)$ is not complicated, since one has

$$E_k(x) + D_k(x) + \ell_k(x) = x\ell'_k(x), \tag{86}$$

similarly to (55).

Proposition 16. *The average size of the left subtree of the leftist trees in $\mathscr{L}_{k,n}$ is asymptotic to $\lambda_k n$ as $n\to\infty$. Here*

$$\lambda_k \equiv \left(1 - \sum_{j=1}^{k-2}\alpha_j\right)\frac{\rho\ell_{k-1}(\rho)}{\alpha_k}. \tag{87}$$

Proof. One can either compute the asymptotic behaviour of $E_{k,n}/\ell_{k,n}$, for large n and fixed k, or else evaluate $\lim_{x\to\rho^-} E_k(x)/(x\ell_k'(x))$. $\square$

The numerical values of the first few terms of the sequence $\{\lambda_k\}$ – which are shown in Table 2 – suggest that this sequence converges to 1/2. This can be easily confirmed with the help of (42) and Theorem 2, and it means that on average the left and the right subtrees become more and more balanced as the d-number increases (i.e., after taking $n\to\infty$).

The average size of the left subtree of the leftist trees in $\mathscr{L}_n$ is asymptotic to λn, with λ being the constant defined by

$$\lambda \equiv \sum_{k=1}^{\infty} \lambda_k \alpha_k = \ell_1(\rho) - \rho \sum_{k=2}^{\infty} \sum_{j=1}^{k-1} \ell_k(\rho)\alpha_j = 0.75445279\ldots \tag{88}$$

(which has also been evaluated numerically in [9]).

We could now compute e.g. the average number of nodes in the left subtree of the left subtree of the leftist trees in $\mathscr{L}_n$. The numerical value we have obtained is different from λ^2, which is in contradiction with a depth independent branching ratio.

6. The average height of leftist trees

The average height is not an additive weight, since linearity is lacking: if we try to write a recursive relation based on Fig. 3 we will find a troublesome *max* operator, or something equivalent. That is why the method developed by Flajolet and Odlyzko [4] for determining the average height of a sizable class of simple (or simply generated) families of trees will be followed closely. However, leftist trees are not simply generated, and some modifications are required.

The generating functions (polynomials, really) for trees with n nodes are height *less or equal to h* – herein denoted by $\ell^{[h]}(x)$ – are among the basic objects considered in [4]. The present analysis will have to include the polynomials $\ell_k^{[h]}(x)$ (the analogue of $\ell^{[h]}(x)$ but for a fixed d-number k). In analogy with (3) one can easily show that

$$\ell_k^{[h]}(x) = x\ell_{k-1}^{[h-1]}(x) \sum_{j\geqslant k-1} \ell_j^{[h-1]}(x), \quad k>0, \tag{89}$$

while $\ell_0^{[h]}(x)=1$. Summation over k leads to

$$\ell^{[h]}(x) = 1 + \frac{x}{2}(\ell^{[h-1]}(x))^2 + \frac{x}{2}\sum_k (\ell_k^{[h-1]}(x))^2. \tag{90}$$

Alas – and this is the main problem – we found no closed recurrence for $\ell^{[h]}(x)$ (that is, in terms of $\ell^{[h-1]}(x)$ only). Notwithstanding, it is still possible to proceed; just like in the case of simple families of trees, the differences $e^{[h]}(x)=\ell(x)-\ell^{[h]}(x)$ still follow an asymptotic recurrence of the form

$$e^{[h]}(x) \sim \sigma_1(x)e^{[h-1]}(x) + \sigma_2(x)(e^{[h-1]}(x))^2 + \cdots. \tag{91}$$

As for the differences $e_k^{[h]}(x) = \ell_k(x) - \ell_k^{[h]}(x)$ we will focus our attention on an asymptotic relation of the type

$$e_k^{[h]}(x) \sim \sigma_{k,1}(x) e^{[h-1]}(x) + \sigma_{k,2}(x)(e^{[h-1]}(x))^2 + \cdots, \tag{92}$$

which is more useful than a recursive one. For $\ell_1^{[h]}(x)$ this may be checked immediately; in this case the expansion is exact and finite.

$$\ell_1(x) - \ell_1^{[h]}(x) = x(\ell(x) - \ell^{[h-1]}(x)) : \tag{93}$$

Thus, $\sigma_{1,1}(x) = x$ and $\sigma_{1,i}(x) = 0$ for $i > 1$ (note also that $\sigma_{0,i}(x) = 0$ for every i). Using recursion over k it is easily seen that $e_k^{[h]}(x)$ contains a piece proportional to $e^{[h-1]}(x)$, but a direct analysis is challenging because there are other terms proportional to $e^{[h-2]}(x)$, etc. Hence, to determine the functions $\sigma_i(x)$ and $\sigma_{k,i}(x)$ we simply insert the asymptotic expansions (91) and (92) into the identity

$$\begin{aligned}\ell(x) - \ell^{[h]}(x) = {} & x\ell(x)(\ell(x) - \ell^{[h-1]}(x)) + x\sum_{k=0}^{\infty} \ell_k(x)(\ell_k(x) - \ell_k^{[h-1]}(x)) \\ & - \frac{x}{2}(\ell(x) - \ell^{[h-1]}(x))^2 - \frac{x}{2}\sum_{k=0}^{\infty}(\ell_k(x) - \ell_k^{[h-1]}(x))^2,\end{aligned} \tag{94}$$

which follows from (90). To first order, and for $k > 0$, this gives

$$\begin{aligned}\sigma_{k,1}(x)\sigma_1(x) = {} & x\sigma_{k-1,1}(x)\left(\ell(x) - \sum_{j=0}^{k-2}\ell_j(x)\right) \\ & + x\ell_{k-1}(x)\left(\sigma_1(x) - \sum_{j=0}^{k-2}\sigma_{j,1}(x)\right).\end{aligned} \tag{95}$$

If one now recalls the properties of the numbers α_k one will realize that

$$\sigma_1(\rho) = 1, \quad \sigma_{k,1}(\rho) = \alpha_k \tag{96}$$

is a solution of the above recurrence for $x = \rho$ (but finding a solution for arbitrary x is another matter). Here we note as well that there cannot be more than one solution (by now we can already exclude the null solution).

Working out the expansion to second order (and taking $x = \rho$, since terms of higher order are not needed) leads to

$$\begin{aligned}\sigma_{k,2}(\rho) = {} & \rho\sigma_{k-1,2}(\rho)\left(\ell(\rho) - \sum_{j=0}^{k-2}\ell_j(\rho)\right) + \rho\ell_{k-1}(\rho)\left(\sigma_2(\rho) - \sum_{j=0}^{k-2}\sigma_{j,2}(\rho)\right) \\ & - \rho\sigma_{k-1,1}(\rho)\left(1 - \sum_{j=0}^{k-2}\sigma_{j,1}(\rho)\right) - \sigma_{k,1}(\rho)\sigma_2(\rho), \quad k > 0\end{aligned} \tag{97}$$

(and we know that $\sigma_{0,2}(x) = 0$). This recurrence will help in finding the solution of

$$\sigma_2(\rho) - \sum_{k=0}^{\infty}\sigma_{k,2}(\rho) = 0. \tag{98}$$

There cannot be more than one solution, since every $\sigma_{k,2}(\rho)$ is linear in $\sigma_2(\rho)$ (if evaluated recursively by means of (97), that is). With the help of a computer one may find $\sigma_2(\rho) = -0.22597287\ldots$ (in the notation of [4], $\sigma_2(\rho)$ correspond to $-1/c_2$).

What is left now is to analyze $e^{[h]}(x)$ for x not necessarily equal to ρ (but close enough to ρ). We will find that there is again a strong similarity with the results obtained for simple families of trees [4]. In the first place we assume that similarity – which is by no means surprising, given all the results obtained so far, including the expansion of $\ell(x)$ near $x = \rho$ – and write

$$\begin{aligned} \sigma_1(x) &= \sigma_1(\rho) + \zeta A(\rho - x)^{1/2} + \cdots, \\ \sigma_{k,1}(x) &= \sigma_{k,1}(\rho) + \zeta_k A(\rho - x)^{1/2} + \cdots, \end{aligned} \tag{99}$$

where A denotes as the multiplicative constant that appears in the expansion (38) of $\ell(x)$:

$$A = -\left(\frac{2}{\eta_2 \rho^2}\right)^{1/2}. \tag{100}$$

Inserting the above expansions into (95) gives

$$\begin{aligned} \zeta_k = \rho\zeta_{k-1}\left(\ell(\rho) - \sum_{j=0}^{k-2} \ell_j(\rho)\right) + \rho\ell_{k-1}(\rho)\left(\zeta - \sum_{j=0}^{k-2}\zeta_j\right) \\ -\alpha_k\zeta + 2\rho\alpha_{k-1}\left(1 - \sum_{j=0}^{k-2}\alpha_j\right), \quad k > 0. \end{aligned} \tag{101}$$

The solution for ζ (and simultaneously for $\zeta_1, \zeta_2, \ldots$) may be found in a manner akin to method used to find $\sigma_2(\rho)$. In the case the equation is

$$\zeta - \sum_{k=0}^{\infty} \zeta_k = 0, \tag{102}$$

and the numerical solution is $\zeta = 0.45194574\ldots$ (the main point is that it is nonzero). It may now be seen why there cannot be terms in (99) that are dominant with respect to $(\rho - x)^{1/2}$, such as $(\rho - x)^{1/4}$. In that case one would obtain a recurrence almost identical to (101) but for the last term (the only term independent of ζ and ζ_k, and which comes from the expansion of $\ell(x)$ and/or $\ell_k(x)$), which would be absent. Then all the ζ_k would be proportional to ζ and (102) would reduce to $\zeta = 0$.

By piecing together the partial results obtained so far one is lead to the following theorem.

Theorem 3. *The average height of leftist trees with n nodes is asymptotic to*

$$\xi_2\sqrt{\pi n} \tag{103}$$

as $n \to \infty$, with

$$\xi_2 = \frac{\rho^2}{|\sigma_2(\rho)|}\left(\frac{\ell_1''(\rho\ell_1(\rho))}{2(\ell_1(\rho) + \rho)}\right)^{1/2} = 1.81349371\ldots\,. \tag{104}$$

Hence, the asymptotic average height of leftist trees is *less* than the average height of binary trees with the same number of nodes (which is given by $2\sqrt{\pi n}$ – see [4]), at least for sufficiently large n; as it is known, a similar relation holds for the average node depth. Rather conspicuously, the numerical value of ξ_2 looks very much like $2\xi_1$; this equality holds numerically for over 1000 decimal digits, and so we conjecture that it is exact.

7. Final remarks and acknowledgments

This paper has benefited from various suggestions made by the referees, as well as from some of the references brought to my attention. For the sake of accuracy a brief report of the latter contributions is included below; whenever appropriate, some comments about those references are also added. Since those events spanned over a long period of time, an approximate date (the year I got the information) is associated with each event.

Ref. [8] was kindly pointed out by D. E. Knuth (1995). During the revision of this paper, Prof. Knuth informed me (1998) that Eq. (6) had been written down by his former student Luis Trabb Pardo somewhere in the period 1978–1980, but was left unpublished until very recently, due to lack of further progress (in the meantime, however, he communicated that result to a small number of people). There is now a new edition of Ref. [13], dated 1998, that contains some updates on leftist trees, as well on alternative combinatorial structures that may supersede them.

Ref. [9] (1997), and later (2000) Refs. [3,11] and the new edition of [13] were pointed out by one of the referees. The existence of Ref. [9] led to a major revision of the original manuscript. A number of results – mainly Theorems 2 and 3 (as well as the whole of Section 6), and the results of Section 4 (that is, in their present generalized form) – were included only on the first revised version (1999).

In [11] a possible generalization of leftist trees, called *leftist simply generated trees* (this is actually an unfortunate designation), was introduced and discussed. One of the results presented therein shows that the functional equation satisfied by that class of trees is a simple generalization of (6), while its proof shows that the proof of Theorem 1 (in this paper) can be made purely combinatorial. The asymptotic number of those trees [3] is actually quite similar to that of leftist trees (there is now a reasonable expectation that their average height is also proportional to $n^{1/2}$).

However, there are also some potentially misleading statements in [11] that ought not to be left unnoticed. One of them is about the functional equation (6) being *well known* for more than 20 years; here a quick look at Refs. [8,9] will clarify the matter. Another such point involves the definition of simply generated trees, which had its root in [14]; although the examples treated in [11] are indeed families of that kind, the definition used is not entirely correct (this same imperfect definition also appeared in [10]). The functional equation for the generating function is not a sufficient condition, there are also rules for the recursive construction of simple trees. Due to lack of space, we will not discuss this issue in detail; let us just mention that a comprehensive analysis would have to include a discussion of [14], which also contains some (very) minor problems.

Last but not least, I would like to acknowledge the hospitality of the Dutch Institute for Nuclear and High Energy Physics (NIKHEF) in Amsterdam, where the very first part of this work was done.

Appendix A. The numerical calculations

This appendix is intended to explain how the numerical results stated in Proposition 8 were obtained. The power series for $\ell_1(\rho)$ is expected to converge quite slowly, since ρ itself defines the circle of convergence. Thus it should not be surprising that one tries to transform an equation that depends on the evaluation of $\ell_1(x)$ at $x=\rho$ into another one that does not require that evaluation to be done at points lying close to the border of the circle of convergence

Proposition A.1. *For any x_0 such that $0 < x_0 \leqslant \rho$ define $x_1 = x_0\ell_1(x_0)$. Then*

$$x_0 = \frac{2x_1}{\ell_1(x_1) + \sqrt{4x_1 + \ell_1^2(x_1)}}. \tag{A.1}$$

Proof. Substituting x_0 for x in (6) and multiplying by x_0 gives the quadratic equation $(x_0)^2 + \ell_1(x_1)x_0 - x_1 = 0$. The only positive solution for x_0 is given by (A.1). □

Let $\{x_n\}$ be the sequence of positive numbers defined in (76), which satisfies the recurrence

$$x_n = x_{n-1}\ell_1(x_{n-1}). \tag{A.2}$$

The limit of that sequence is equal to zero since $\sum_{n=0}^{\infty} x_n$ is easily seen to converge. Now let us imagine we are given (for some k larger than 1) a very good numerical approximation to $x_k, \ell_1(x_k), \ell_1'(x_k)$, and $\ell_1''(x_k)$; then from this we may easily derive approximations for $\rho, \ell_1(\rho), \ell_1(x_1), \ell_1'(x_1)$, and $\ell_1''(x_1)$. Here is how: let

$$y_k = \ell_1(x_k), \quad y_k' = \ell_1'(x_k), \quad y_k'' = \ell_1''(x_k), \tag{A.3}$$

and then just iterate, in this order,

$$\begin{aligned} x_n &= \frac{2x_{n+1}}{y_{n+1} + \sqrt{4x_{n+1} + (y_{n+1})^2}}, \\ y_n &= x_n + y_{n+1}, \\ y_n' &= \frac{1 + y_n y_{n+1}'}{1 - x_n y_{n+1}'}, \\ y_n'' &= \frac{(x_n + y_n)^2 y_{n+1}''}{(1 - x_n y_{n+1}')^3} + \frac{2y_n' y_{n+1}'}{1 - x_n y_{n+1}'}, \end{aligned} \tag{A.4}$$

for $n = k-1, k-2, \ldots, 0$ (in fact y_0' and y_0'' should not be evaluated, for obvious reasons). These formulae are just a disguise of (A.1), (6), (20), and (22), respectively.

In practical terms we choose a value for k and make an initial guess for x_k, and then use a numerical algorithm to find the positive solution of

$$x_0 y_1' = 1. \tag{A.5}$$

Any reasonable one-dimensional root-finding algorithm may be used in this search, including the ones that require the evaluation of the function's derivative (e.g., Newton–Raphson method). An approximation to $\ell_1(x)$ is required, and for this we take a truncation of the corresponding power series. Now the idea is that if k is taken large enough then even a single term of the power series may provide the desired accuracy. This is not unreasonable since for small x one has $\ell_1(x) \approx x$, meaning that the x_n converge quadratically to zero i.e., $x_{n+1} \approx (x_n)^2$. The numerical values of the x_k for the first few values of k may be easily computed from the data presented in Table 2. It is actually very easy to find approximations accurate to several hundreds of decimal places. The method just described may be used to solve other equations involving $\ell_1(x)$ and its derivatives, and in that case x_0 would be a root of the equation to be solved.

Appendix B. Some further relations

The next results provide a means to cross-check the accuracy of the numerical calculations (that is in fact the main reason for their inclusion in this paper). Some of the identities involve constants and can be used directly, but some of them are more general; in the latter case replacing x by a convenient constant (e.g., ρ) gives an identity of the former type. It is also possible that some of these identities may be used to simplify the expressions for expected average weights not considered herein (it is certainly true that in a few of the cases considered in this paper we have observed some unexpected simplifications).

The first identity is closely related to the one proved in [8, Lemma 2.1]:

$$x \sum_{k=0}^{\infty} \ell_k^2(x) = 2\ell(x) - x\ell^2(x) - 2. \tag{B.1}$$

It may be easily proved by summing (3) for k from 1 to ∞ and performing a few simple manipulations. Actually one can go one level higher and (using (3) within itself) prove that

$$x^2 \sum_{k=0}^{\infty} \ell_k^3(x) = x^2\ell^3(x) - 3(1-x)\ell(x) + 3, \tag{B.2}$$

but for higher degrees such a simplification does not seem to occur. Other identities follow by differentiation of the latter two formulae with respect to x, with the first order derivatives leading to

$$\rho \sum_{k=1}^{\infty} \alpha_k \ell_k(\rho) = 1 - \ell_1(\rho) \tag{B.3}$$

and

$$\rho^2 \sum_{k=1}^{\infty} \alpha_k \ell_k(\rho)^2 = \ell_1(\rho)^2 + \rho - 1. \tag{B.4}$$

Another way of obtaining identities is through the generic recurrence

$$U_k(x) = C(x) \left(U_{k-1}(x) \left(\ell(x) - \sum_{j=0}^{k-2} \ell_k(x) \right) + \ell_{k-1}(x) \left(U(x) - \sum_{j=0}^{k-2} U_j(x) \right) \right) + Y_k(x). \tag{B.5}$$

If the $U_k(x)$ satisfy the previous recurrence, with $U(x)$ being the sum of the series with general term $U_k(x)$ (here we assume that all the involved infinite series are convergent) then a trivial summation gives

$$\sum_{k=0}^{\infty} U_k(x) = C(x) \left(\ell(x) \cdot U(x) + \sum_{k=0}^{\infty} U_k(x) \ell_k(x) \right) + \sum_{k=0}^{\infty} Y_k(x). \tag{B.6}$$

For example, (B.5) reproduces (40) if $C(x) = x$, $x = \rho$, $Y_k(x) = 0$, and $U_k(\rho) = \alpha_k$; then it follows that (B.6) also leads to (B.3). On the other hand, applying this procedure to e.g. (97) produces a new identity.

References

[1] E.A. Bender, Asymptotic methods in enumeration, SIAM Rev. 16 (1974) 485–515.
[2] C.A. Crane, Linear lists and priority queues as balanced binary trees, Technical Report STAN-CS-72-259, Stanford University, 1972.
[3] M. Drmota, The asymptotic number of leftist trees, preprint, Technical University of Vienna, 1999.
[4] P. Flajolet, A. Odlyzko, The average height of binary trees and other simple trees, J. Comput. System Sci. 25 (1982) 171–213.
[5] G.W. Ford, G.E. Uhlenbeck, Combinatorial problems in the theory of graphs, III, Proc. Natl. Acad. Sci. U.S.A. 42 (1956) 529–535.
[6] F. Harary, E.M. Palmer, Graphical Enumeration, Academic Press, New York, 1973.
[7] F. Harary, R.W. Robinson, A.J. Schwenk, Twenty-step algorithm for determining the asymptotic number of trees of various species, J. Austral. Math. Soc. 20 (Series A) (1975) 483–503.
[8] R. Kemp, A note on the number of leftist trees, Inform. Process. Lett. 25 (1987) 227–232.
[9] R. Kemp, Further results on leftist trees, in: M. Karoński, J. Jaworski, A. Ruciński (Eds.), Random Graphs '87, Wiley, New York, 1990, pp. 103–130.
[10] R. Kemp, The expected additive weight of trees, Acta Inform. 26 (1989) 711–740.
[11] R. Kemp, On leftist simply generated trees, J. Automat. Languages Combin. 4 (1999) 313–331.
[12] D. Knuth, The Art of Computer Programming, Vol. 1, Addison-Wesley, Reading, MA, 1973.
[13] D. Knuth, The Art of Computer Programming, Vol. 3, Addison-Wesley, Reading, MA, 1973.
[14] A. Meir, J.W. Moon, On the altitude of nodes in random trees, Can. J. Math. 30 (1978) 997–1015.
[15] A.M. Odlyzko, Asymptotic enumeration methods, in: R.L. Graham, M. Grötschel, L. Lovász (Eds.), Handbook of Combinatorics, Elsevier, Amsterdam, 1995, pp. 1063–1229.
[16] R. Otter, The number of trees, Ann. Math. 49 (1948) 583–599.
[17] G. Pólya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und Chemisch Verbindungen, Acta Math. 68 (1937) 145–254.
[18] G. Pólya, R.C. Read, Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds, Springer, New York, 1987.